Biomarker development: current issues and perspectives

Raoul Belzeaux, Julia-Lou Consoloni, El Chérif Ibrahim

Biomarker development: current issues and perspectives

Raoul Belzeaux[1,2,3,4], Julia-lou Consoloni[1,2,3], El Chérif Ibrahim[1,3]

[1]Aix-Marseille Univ, CNRS, CRN2M-UMR7286, Marseille, France

[2]Pôle de Psychiatrie, Assistance Publique-Hôpitaux de Marseille, Marseille, France

[3]Fondation FondaMental, Fondation de Recherche et de Soins en Santé Mentale, Créteil, France

[4]McGill Group for Suicide Studies, Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada

Corresponding author:

El Chérif Ibrahim, 51 Bd Pierre Dramard, 13344 Marseille Cedex 15 France

Tel: +33 (0)4 91 69 89 56; Fax: +33 (0)4 91 69 89 20

Email: el-cherif.ibrahim@univ-amu.fr

To the editor:

We have read the letter of Ho et al. in response to our work with great interest. We agree with their general comment regarding the interest of evaluating agreement between two multiplex immunoassays.

Ho et al. reminds us that correlation analysis suffers from limitations in assessing agreement between methods. Such agreement should be assessed using Bland-Altman plots. This method is largely known as the original 1986 paper of Bland and Altman has been the most frequently cited article ever to appear in the *Lancet* and is one of the ten most frequently cited statistical articles ever (Bland and Altman, 2012).

Nevertheless, we decided to Spearman's rank correlation, since our original intent was to compare methods and biological fluids, which is not exactly the same as assessing the degree of agreement. Ultimately, several conditions are required to evaluate agreement with Bland-Altman graphics, Differences of measurements between methods need to be normally distributed and the variability of the paired differences has to be uniform along the range of measurements (homoscedasticity). Heteroscedastic data should be transformed logarithmically or investigated with an analysis based on ranks (Atkinson and Nevill, 1998). Furthermore, the Bland-Altman method is based on a qualitative appreciation of concordance. Deciding whether the agreement between methods or samples is sufficient depends on the context in which the measurements are used (Bartlett and Frost, 2008). In psychiatry, the frequent absence of a gold standard flow chart for measuring of soluble proteins and the lack of tools providing the right absolute amount of the assayed protein renders any decision on appropriate limits of agreement for evaluating two relative quantification assays highly questionable.

Ho et al. also recommend the use of Bland-Altman plots to evaluate the stability of cytokine measurements in healthy individuals over time. They also state that the chance that cytokine levels remain constant over a 210-day period is unlikely. However, a biomarker that would discriminate psychiatric patients from the general population has to demonstrate both a higher inter-individual variability and a lower intra-individual variability in healthy subjects. To explore stability within subjects, we chose to calculate the Intra-Class Correlation coefficient (ICC) in our analyses, which is a well-recommended method (Liu et al., 2016). With respect to the aforementioned remark on the required stability of a biomarker over a long time period in healthy individuals, we do not see any advantage of segregating the calculation of ICC over different retest intervals.

Too many studies describing biomarkers only compared differences between groups without clear validation process of measurement for reproducibility and stability in healthy subjects. Correlations could be a first step to seek for disagreement, while other more specific analyses should be used when available technical tools reach consistency and provide validated measurements compared to absolute values.

References

Atkinson, G., Nevill, A.M., 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26, 217-238.

Bartlett, J.W., Frost, C., 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 31, 466-475.

Bland, J.M., Altman, D.G., 2012. Agreed statistics: measurement method comparison. Anesthesiology 116, 182-185.

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., Tu, X.M., 2016. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry 28, 115-120.